Sebastian Pütz: Insights into subword embeddings

Word embeddings with subword representations are a staple choice in NLP, but little has been done to explore how these models store information. In this talk, I will show that fastText-style models often leave a large partition of their vector space untrained and discuss how an explicit ngram lookup-table can alleviate this issue. Moreover, I will provide insights into how downstream tasks are influenced by changing the subword lookup.